


[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☐ The Guide

THE ACM DIGITAL LIBRARY

[Feedback](#)

correcting documents multiple references Found
 Terms used: 481 of
 correcting documents multiple references 244,119

 Sort
 results
 by

 Display
 results


[Save](#) Refine

[results](#) these
[to a](#) results
[Binder](#) with
[Advanced](#)

 Open
 results
 in a new
 window
[Search](#)
 Try this
 search
 in [The](#)
[ACM](#)
[Guide](#)

 Results 1 - 20 of 481 Result page: 1 2 3 4 5 6 7 8 9 10 [next](#)
[>>](#)

1 [Collective entity resolution in relational data](#)



Indrajit Bhattacharya, Lise Getoor

 March ACM Transactions on Knowledge Discovery from Data (TKDD),
 2007 Volume 1 Issue 1

Publisher: ACM

 Full text available: [pdf\(511.57 KB\)](#)

 Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)
Bibliometrics: Downloads (6 Weeks): 40, Downloads (12 Months): 795, Citation Count: 0

Many databases contain uncertain and imprecise references to real-world entities. The absence of identifiers for the underlying entities often results in a database which contains multiple references to the same entity. This can lead not only to data ...


Keywords: Entity resolution, data cleaning, graph clustering, record linkage

2 [Jump: a system for interactive, tangible queries of paper](#)



Michael Terry, Janet Cheung, Justin Lee, Terry Park, Nigel Williams
May 2007 GI '07: Proceedings of Graphics Interface 2007

Publisher: ACM

Full text available:  [pdf\(3.63 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 6, Downloads (12 Months): 147, Citation Count: 0

This paper introduces *Jump*, a prototype computer vision-based system that transforms paper-based architectural documents into *tangible query interfaces*. Specifically, *Jump* allows a user to obtain additional information related to a given ...

Keywords: augmented reality, error recovery, recognition error, tangible user interface, zoomable user interface

3 [Automatic construction of multifaceted browsing interfaces](#)



Wisam Dakka, Panagiotis G. Ipeirotis, Kenneth R. Wood
October CIKM '05: Proceedings of the 14th ACM international conference on
2005 Information and knowledge management

Publisher: ACM

Full text available:  [pdf\(397.60 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 11, Downloads (12 Months): 122, Citation Count: 0

Databases of text and text-annotated data constitute a significant fraction of the information available in electronic form. Searching and browsing are the typical ways that users locate items of interest in such databases. Interfaces that use multifaceted ...


Keywords: browsing, faceted classification, faceted navigation, hierarchy construction, multifaceted hierarchies

4 [Confidence estimation for NLP applications](#)



Simona Gandrabur, George Foster, Guy Lapalme
October ACM Transactions on Speech and Language Processing (TSLP),
2006 Volume 3 Issue 3

Publisher: ACM

Full text available:  [pdf\(1.17 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 12, Downloads (12 Months): 156, Citation Count: 0

Confidence measures are a practical solution for improving the usefulness of Natural Language Processing applications. Confidence estimation is a generic machine learning approach for deriving confidence measures. We give an overview of the application ...


Keywords: Confidence measures, computer-aided machine translation, confidence estimation, speech recognition, spoken language understanding, statistical machine learning

5 [Towards a query optimizer for text-centric tasks](#)



Panagiotis G. Ipeirotis, Eugene Agichtein, Pranay Jain, Luis Gravano
November ACM Transactions on Database Systems (TODS), Volume 32 Issue 4
2007

Publisher: ACM

Full text available:  [pdf\(1.10 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 41, Downloads (12 Months): 208, Citation Count: 0

Text is ubiquitous and, not surprisingly, many important applications rely on textual data for a variety of tasks. As a notable example, information extraction applications derive structured relations from unstructured text; as another example, focused ...


Keywords: Metasearching, distributed information retrieval, focused crawling, information extraction, text database selection

6 [Compactly encoding unstructured inputs with differential compression](#)



Miklos Ajtai, Randal Burns, Ronald Fagin, Darrell D. E. Long, Larry Stockmeyer
May 2002 Journal of the ACM (JACM), Volume 49 Issue 3

Publisher: ACM

Full text available:  [pdf\(348.32 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 14, Downloads (12 Months): 83, Citation Count: 8

The subject of this article is *differential compression*, the algorithmic task of finding common strings between versions of data and using them to encode one version compactly by describing it as a set of changes from its companion. A main goal ...

Keywords: Delta compression, differencing, differential compression

7 [RCV1: A New Benchmark Collection for Text Categorization Research](#)

David D. Lewis, Yiming Yang, Tony G. Rose, Fan Li

December 2004 The Journal of Machine Learning Research, Volume 5

Publisher: MIT Press

Full text available:  [pdf\(628.29 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#),
[index terms](#), [review](#)

Bibliometrics: Downloads (6 Weeks): 22, Downloads (12 Months): 204, Citation Count: 54

Reuters Corpus Volume I (RCV1) is an archive of over 800,000 manually categorized newswire stories recently made available by Reuters, Ltd. for research purposes. Use of this data for research on text categorization requires a detailed understanding ...


8 [Integrating XML data sources using approximate joins](#)



Sudipto Guha, H. V. Jagadish, Nick Koudas, Divesh Srivastava, Ting Yu

March 2006 ACM Transactions on Database Systems (TODS), Volume 31 Issue 1

Publisher: ACM

Full text available:  [pdf\(1.39 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 23, Downloads (12 Months): 186, Citation Count: 1

XML is widely recognized as the data interchange standard of tomorrow because of its ability to represent data from a variety of sources. Hence, XML is likely to be the format through which data from multiple sources is integrated. In this article, we ...

Keywords: Data integration, XML, approximate joins, joins, tree edit distance


9 [Projectors: advanced graphics and vision techniques](#)



Ramesh Raskar

August 2004 SIGGRAPH '04: ACM SIGGRAPH 2004 Course Notes

Publisher: ACM

Full text available:  [pdf\(6.53 MB\)](#)

Additional Information: [full citation](#), [cited by](#)

Bibliometrics: Downloads (6 Weeks): 45, Downloads (12 Months): 658, Citation Count: 1

10 [Multiple selections in smart text editing](#)



Robert C. Miller, Brad A. Myers

January 2002 IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces

Publisher: ACM

Full text available: [pdf\(150.05 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#),

[index terms](#)

Bibliometrics: Downloads (6 Weeks): 6, Downloads (12 Months): 28, Citation Count: 3

Multiple selections, though heavily used in file managers and drawing editors, are virtually nonexistent in text editing. This paper describes how multiple selections can automate repetitive text editing. *Selection guessing* infers a multiple selection ...

Keywords: LAPIS, PBD, automated text editing, pattern matching, programming-by-demonstration, search-and-replace

11 [Visual signature based identification of Low-resolution document images](#)



Ardhendu Behera, Denis Lalanne, Rolf Ingold

October 2004 DocEng '04: Proceedings of the 2004 ACM symposium on Document engineering

Publisher: ACM

Full text available: [pdf\(2.00 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index](#)

[terms](#)

Bibliometrics: Downloads (6 Weeks): 6, Downloads (12 Months): 70, Citation Count: 1

In this paper, we present (a) a method for identifying documents captured from low-resolution devices such as web-cams, digital cameras or mobile phones and (b) a technique for extracting their textual content without performing OCR. The first method ...

Keywords: document visual signature, document-based meeting retrieval, documents' content extraction, low-resolution document image identification

12 [Trace-driven memory simulation: a survey](#)



Richard A. Uhlig, Trevor N. Mudge

June 1997 ACM Computing Surveys (CSUR), Volume 29 Issue 2

Publisher: ACM

Full text available: [pdf\(636.11 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#),

[index terms](#), [review](#)

Bibliometrics: Downloads (6 Weeks): 15, Downloads (12 Months): 200, Citation Count: 40

As the gap between processor and memory speeds continues to widen, methods for evaluating memory system designs before they are implemented in hardware are becoming increasingly important. One such method, trace-driven memory simulation, has been the ...

Keywords: TLBs, caches, memory management, memory simulation, trace-driven simulation

13 [Dynamic spatial approximation trees](#)



Gonzalo Navarro, Nora Reyes

August 2007 Journal of Experimental Algorithmics (JEA), Volume 12

Publisher: ACM

Full text available: [pdf\(2.33 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 31, Downloads (12 Months): 95, Citation Count: 0

Metric space searching is an emerging technique to address the problem of efficient similarity searching in many applications, including multimedia databases and other repositories handling complex objects. Although promising, the metric space approach ...

Keywords: Multimedia databases, similarity or proximity search, spatial and multidimensional search, spatial approximation tree

14 [Summarization of noisy documents: a pilot study](#)

Hongyan Jing, Daniel Lopresti, Chilin Shih

May Proceedings of the HLT-NAACL 03 on Text summarization workshop - 2003 Volume 5, Volume 5

Publisher: Association for Computational Linguistics

Full text available: [pdf\(107.21 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#)

Bibliometrics: Downloads (6 Weeks): 0, Downloads (12 Months): 28, Citation Count: 0

We investigate the problem of summarizing text documents that contain errors as a result of optical character recognition. Each stage in the process is tested, the error effects analyzed, and possible solutions suggested. Our experimental results show ...

15 [CiteSeer: a scalable autonomous scientific digital library](#)



Huajing Li, Isaac G. Councill, Levent Bolelli, Ding Zhou, Yang Song, Wang-Chien Lee, Anand Sivasubramaniam, C. Lee Giles
May InfoScale '06: Proceedings of the 1st international conference on Scalable 2006 information systems

Publisher: ACM

Full text available: [pdf\(229.57 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 4, Downloads (12 Months): 103, Citation Count: 0

CiteSeer is a scientific literature digital library and search engine which automatically crawls and indexes scientific documents in the fields of computer and information science. Since it's inception in 1997 CiteSeer has grown to index over 730,000 ...

16 [Electronic document addressing: dealing with change](#)



Helen Ashman
September 2000 ACM Computing Surveys (CSUR), Volume 32 Issue 3
Publisher: ACM

Full text available: [pdf\(92.20 KB\)](#)

Additional Information: [full citation](#), [appendices and supplements](#), [abstract](#), [references](#), [cited by](#), [index terms](#), [review](#)

Bibliometrics: Downloads (6 Weeks): 10, Downloads (12 Months): 123, Citation Count: 5

The management of electronic document collections is fundamentally different from the management of paper documents. The ephemeral nature of some electronic documents means that the document address (i.e., reference details of the document) can become ...

Keywords: 404, link, link integrity

17 [A semidiscrete matrix decomposition for latent semantic indexing information retrieval](#)



Tamara G. Kolda, Dianne P. O'Leary
October 1998 ACM Transactions on Information Systems (TOIS), Volume 16 Issue 4
Publisher: ACM

Full text available: [pdf\(201.61 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#), [review](#)

Bibliometrics: Downloads (6 Weeks): 9, Downloads (12 Months): 112, Citation Count: 20

The vast amount of textual information available today is useless unless it can be effectively and efficiently searched. The goal in information retrieval is to find documents that are relevant to a given user query. We can represent and document collection ...

Keywords: data mining, latent semantic indexing, semidiscrete decomposition, singular-value decomposition, text retrieval

18 A fast orientation and skew detection algorithm for monochromatic document




images

Bruno Tenório Ávila, Rafael Dueire Lins

November 2005 DocEng '05: Proceedings of the 2005 ACM symposium on Document engineering

Publisher: ACM

Full text available:  [pdf\(509.46 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 2, Downloads (12 Months): 143, Citation Count: 2

Very often in the digitization process, documents are either not placed with the correct orientation or are rotated of small angles in relation to the original image axis. These factors make more difficult the visualization of images by human users, ...

Keywords: monochromatic document image, orientation and skew detection


19 Metadata handling: A video perspective



Chitra L. Madhwacharyula, Marc Davis, Philippe Mulhem, Mohan S. Kankanhalli

November 2006 ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), Volume 2 Issue 4

Publisher: ACM

Full text available:  [pdf\(9.35 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 19, Downloads (12 Months): 200, Citation Count: 0

This article addresses the problem of processing the annotations of preexisting video productions to enable reuse and repurposing of metadata. We introduce the concept of automatic content-based editing of preexisting semantic home video metadata. We ...

Keywords: Preexisting video, content, editing, metadata, representation, reuse, semantic

20 [Using data mining techniques and rough set theory for language modeling](#)



Yong Chen, Kwok-Ping Chan

April 2007 ACM Transactions on Asian Language Information Processing (TALIP), Volume 6 Issue 1

Publisher: ACM

Full text available: [pdf1858.52](#)



Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 7, Downloads (12 Months): 275, Citation Count: 0

In this article, we propose a new postprocessing strategy, word suggestion, based on a multiple word trigger-pair language model for Chinese character recognizers. With the word suggestion strategy, Chinese character recognizers may even achieve a recognition ...

Keywords: Chinese character recognizer, postprocessing

Results 1 - 20 of 481 Result page: 1 2 3 4 5 6 7 8 9 10 [next](#)



The ACM

Portal is published by the Association for Computing Machinery. Copyright © 2008 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads: [Adobe Acrobat](#)



[QuickTime](#)



[Windows Media Player](#)



[Real Player](#)